EGT2
ENGINEERING TRIPOS PART IIA

Wednesday 2 May 2018 9.30 to 11:10

Module 3M1

MATHEMATICAL METHODS

*Answer not more than **three** questions.*

All questions carry the same number of marks.

*The **approximate** percentage of marks allocated to each part of a question is indicated in the right margin.*

*Write your candidate number __**not**__ your name on the cover sheet.*

STATIONERY REQUIREMENTS
Single-sided script paper

SPECIAL REQUIREMENTS TO BE SUPPLIED FOR THIS EXAM
CUED approved calculator allowed
Engineering Data Book
3M1 Optimization Data Sheet (4 pages)

10 minutes reading time is allowed for this paper at the start of the exam.

You may not start to read the questions printed on the subsequent pages of this question paper until instructed to do so.

1 (a) For a $m \times n$ matrix $\boldsymbol{A}$, consider a singular value decomposition $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$.

(i) What is the relationship between $\boldsymbol{U}$ and $\boldsymbol{V}$ when $\boldsymbol{A}$ is symmetric? [10%]

(ii) In terms of the components of $\boldsymbol{\Sigma}$, when will the computation of a matrix-vector product, $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{c}$, using a computer be susceptible to large round-off errors in terms of the l_2 norm? [10%]

(b) For a symmetric matrix $\boldsymbol{A}$, given a vector $\boldsymbol{x}$ that is close to an eigenvector of $\boldsymbol{A}$ such that $\boldsymbol{A}\boldsymbol{x} \approx \lambda \boldsymbol{x}$ where λ is the associated eigenvalue, prove that an optimal estimate of the eigenvalue is given by

$$\lambda^\star = \frac{\boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}},$$

and give the error measure that is minimised. [30%]

(c) A stationary iterative method for solving $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ involves splitting $\boldsymbol{A}$ such that $\boldsymbol{A} = \boldsymbol{N} - \boldsymbol{P}$ and solving

$$\boldsymbol{x}_{k+1} = \underbrace{\boldsymbol{N}^{-1}\boldsymbol{P}}_{\boldsymbol{M}}\boldsymbol{x}_k + \boldsymbol{N}^{-1}\boldsymbol{b},$$

where the split is chosen such that $\boldsymbol{N}\boldsymbol{x} = \boldsymbol{c}$ is much cheaper to solve than $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{c}$.

Prove the necessary condition to guarantee convergence of this iterative method, and express the condition in terms of a norm of the iteration matrix $\boldsymbol{M}$. [50%]

2 (a) An $N \times N$ transition matrix $\mathbf{P}$ governs transitions in a finite-space homogeneous Markov chain where element p_{jk} yields

$$p_{jk} = P(X_{n+1} = k | X_n = j)$$

(i) Explain why the sum of the elements in any row of $\mathbf{P}$ is equal to 1. [10%]

(ii) By considering the sum of the first $N-1$ columns of $\mathbf{P}-\mathbf{I}$, where $\mathbf{I}$ is the identity matrix, or otherwise, show that $\mathbf{P}$ has at least one eigenvalue equal to 1. [20%]

(iii) By considering the stationary distribution of the process, or otherwise, show that the magnitude of the eigenvalues of the transition matrix must all be less than or equal to 1. [15%]

(b) Fig. 1 shows the state-diagram for surfing three connected web-pages and the associated probability of transitioning between pages.

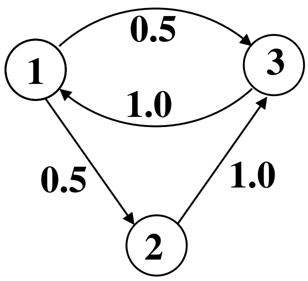


Fig. 1

In addition to the process shown, at each click there is a probability d that an individual simply randomly selects a web-page. Show that the transition matrix for the surfing of an individual is $\tilde{\mathbf{P}}$ where

$$\tilde{\mathbf{P}} = d\mathbf{A} + (1-d)\mathbf{P}$$

where $\mathbf{P}$ is the transition matrix for the process shown in Fig 1.

(i) What are the transition matrices $\mathbf{P}$ and $\mathbf{A}$? [15%]

(ii) Find the stationary distributions for $\tilde{\mathbf{P}}$ when $d = 0$ and $d = 1$. [25%]

(iii) In practice the number of web-pages, N, is very large. Describe an iterative approach for finding the stationary distribution that is efficient as the number of web-pages gets very large. You should compute the complexity of the approach. [15%]

3 The potential energy, V, of a mechanical system is given by the following expression in terms of the free variables x and y,

$$V(x,y) = x^2 + y^2 + 3(x-y)^2 + x - 2y.$$

(a) Starting from $x = 0, y = 0$, execute one step of Steepest Descent minimization, giving the updated location and the function value at that location. [20%]

(b) State what you expect would happen if subsequent steps were taken with the following methods,

(i) Steepest Descent;

(ii) Conjugate Gradients.

[15%]

(c) By considering the Taylor expansion of a function in two dimensions, write down the update rule corresponding to a Newton step. [10%]

(d) Again starting from the $x = 0, y = 0$ position, execute one Newton step and comment on the result. [25%]

(e) State the second order optimality conditions for a local minimum, and verify that the location of the stationary point of the function is indeed a minimum. Calculate the eigenvectors of the Hessian matrix and sketch the contours of the potential energy function $V(x,y)$, also marking the location of the result of part (a). [30%]

4 A factory is manufacturing two kinds of products, A and B, from a similar set of raw materials and similar but not identical processes. The income from selling a unit of A is £2 and for unit of B is £1. The factory manager wishes to maximise the income by making the optimum amount of A and B each day, subject to the following constraints. For one step of the manufacturing, there is only one machine, which can process one unit of either A or B at a time, and it takes 2 hours to process a unit of A and 3 hours to process a unit of B. The machine can be operated for a maximum of 16 hours per day over two shifts. Another step, in which A and B units can be processed together is such that the number of A units processed cannot exceed three plus the number of B units. Due to an agreement with one of the unions, the manager has agreed to make at least half as many A units as he makes of B units.

(a) Convert the above problem to a linear programming optimisation problem, stating the objective function and the constraints. Use variables x_1 and x_2 for the number of units made daily of A and B, respectively. [20%]

(b) Draw the feasible region graphically, and identify the solution by inspection. [20%]

(c) Introduce slack variables to convert the problem to standard form and show that a canonical tableau is given by

$$\begin{pmatrix} 1 & -2 & -1 & 0 & 0 & 0 & 0 \\ 0 & -2 & 1 & 1 & 0 & 0 & 0 \\ 0 & 2 & 3 & 0 & 1 & 0 & 16 \\ 0 & 1 & -1 & 0 & 0 & 1 & 3 \end{pmatrix}$$

[20%]

(d) Execute the simplex method until convergence, starting from the basic feasible point $x_1 = x_2 = 0$ and using the above canonical tableau, showing each pivot element and a new the tableau after each pivot operation. Identify the basic and nonbasic variables at each step by marking the corresponding columns. Indicate the path of the simplex algorithm on your drawing in part (b). [40%]

END OF PAPER

THIS PAGE IS BLANK

3M1

OPTIMIZATION

DATA SHEET

1. Taylor Series Expansion

For one variable:

$$f(x) = f(x^*) + (x - x^*) f'(x^*) + \frac{1}{2}(x - x^*)^2 f''(x^*) + R$$

For several variables:

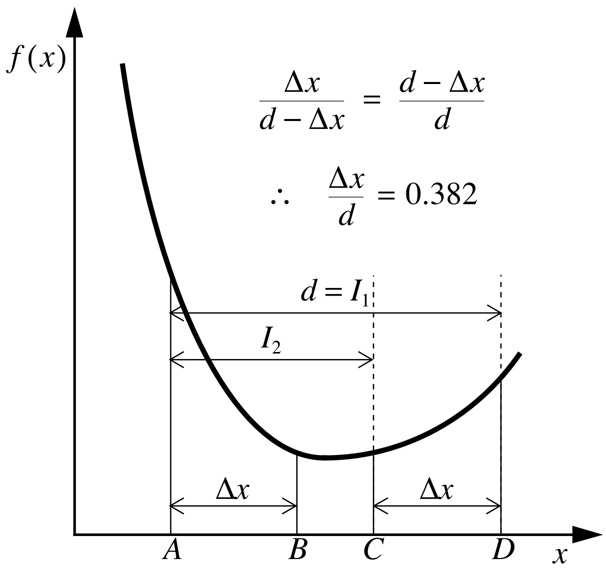
$$f(\mathbf{x}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T \boldsymbol{H}(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) + R$$

where

$$\text{gradient } \nabla f(\mathbf{x}) = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{bmatrix} \quad \text{and hessian } \boldsymbol{H}(\mathbf{x}) = \nabla(\nabla f(\mathbf{x})) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

$\boldsymbol{H}(\mathbf{x}^*)$ is a symmetric $n \times n$ matrix and R includes all higher order terms.

2. Golden Section Method



(a) Evaluate $f(x)$ at points A, B, C and D.

(b) If $f(B) < f(C)$, new interval is $A - C$.

If $f(B) > f(C)$, new interval is $B - D$.
If $f(B) = f(C)$, new interval is either $A - C$ or $B - D$.

(c) Evaluate $f(x)$ at new interior point. If not converged, go to (b).

3. **Newton's Method**

(a) Select starting point $\mathbf{x}_0$

(b) Determine search direction $\mathbf{d}_k = -\boldsymbol{H}(\mathbf{x}_k)^{-1}\nabla f(\mathbf{x}_k)$

(c) Determine new estimate $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{d}_k$

(d) Test for convergence. If not converged, go to step (b)

4. **Steepest Descent Method**

(a) Select starting point $\mathbf{x}_0$

(b) Determine search direction $\mathbf{d}_k = -\nabla f(\mathbf{x}_k)$

(c) Perform line search to determine step size α_k or evaluate $\alpha_k = \dfrac{\mathbf{d}_k^T\mathbf{d}_k}{\mathbf{d}_k^T\boldsymbol{H}(\mathbf{x}_k)\mathbf{d}_k}$

(d) Determine new estimate $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\mathbf{d}_k$

(e) Test for convergence. If not converged, go to step (b)

5. **Conjugate Gradient Method**

(a) Select starting point $\mathbf{x}_0$ and compute $\mathbf{d}_0 = -\nabla f(\mathbf{x}_0)$ and $\alpha_0 = \dfrac{\mathbf{d}_0^T\mathbf{d}_0}{\mathbf{d}_0^T\boldsymbol{H}(\mathbf{x}_0)\,\mathbf{d}_0}$

(b) Determine new estimate $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\mathbf{d}_k$

(c) Evaluate $\nabla f(\mathbf{x}_{k+1})$ and $\beta_k = \left[\dfrac{|\nabla f(\mathbf{x}_{k+1})|}{|\nabla f(\mathbf{x}_k)|}\right]^2$

(d) Determine search direction $\mathbf{d}_{k+1} = -\nabla f(\mathbf{x}_{k+1}) + \beta_k\mathbf{d}_k$

(e) Determine step size $\alpha_{k+1} = -\dfrac{\mathbf{d}_{k+1}^T\nabla f(\mathbf{x}_{k+1})}{\mathbf{d}_{k+1}^T\boldsymbol{H}(\mathbf{x}_{k+1})\,\mathbf{d}_{k+1}}$

(f) Test for convergence. If not converged, go to step (b)

6. **Gauss-Newton Method (for Nonlinear Least Squares)**

If the minimum squared error of residuals $\mathbf{r}(\mathbf{x})$ is sought:

$$\text{Minimise } f(\mathbf{x}) = \sum_{i=1}^{m} r_i^2(\mathbf{x}) = \mathbf{r}(\mathbf{x})^T\mathbf{r}(\mathbf{x})$$

(a) Select starting point $\mathbf{x}_0$

(b) Determine search direction $\mathbf{d}_k = -[\,\boldsymbol{J}(\mathbf{x}_k)^T\boldsymbol{J}(\mathbf{x}_k)\,]^{-1}\,\boldsymbol{J}(\mathbf{x}_k)^T\mathbf{r}(\mathbf{x}_k)$

$$\text{where } \boldsymbol{J}(\mathbf{x}) = \begin{bmatrix} \nabla r_1(\mathbf{x})^T \\ \vdots \\ \nabla r_m(\mathbf{x})^T \end{bmatrix} = \begin{bmatrix} \frac{\partial r_1}{\partial x_1} & \cdots & \frac{\partial r_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial r_m}{\partial x_1} & \cdots & \frac{\partial r_m}{\partial x_n} \end{bmatrix}$$

(c) Determine new estimate $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{d}_k$

(d) Test for convergence. If not converged, go to step (b)

7. **Lagrange Multipliers**

To minimise $f(\mathbf{x})$ subject to m equality constraints $h_i(\mathbf{x}) = 0$, $i = 1, \ldots, m$, solve the system of simultaneous equations

$$\begin{aligned} \nabla f(\mathbf{x}^*) + [\nabla \mathbf{h}(\mathbf{x}^*)]^T \boldsymbol{\lambda} &= 0 \quad (n \text{ equations}) \\ \mathbf{h}(\mathbf{x}^*) &= 0 \quad (m \text{ equations}) \end{aligned}$$

where $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_m]^T$ is the vector of Lagrange multipliers and

$$[\nabla \mathbf{h}(\mathbf{x}^*)]^T = \begin{bmatrix} \nabla h_1(\mathbf{x}^*) & \ldots & \nabla h_m(\mathbf{x}^*) \end{bmatrix} = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_m}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_1}{\partial x_n} & \cdots & \frac{\partial h_m}{\partial x_n} \end{bmatrix}$$

8. **Kuhn-Tucker Multipliers**

To minimise $f(\mathbf{x})$ subject to m equality constraints $h_i(\mathbf{x}) = 0$, $i = 1, \ldots, m$ and p inequality constraints $g_i(\mathbf{x}) \le 0$, $i = 1, \ldots, p$, solve the system of simultaneous equations

$$\begin{aligned} \nabla f(\mathbf{x}^*) + [\nabla \mathbf{h}(\mathbf{x}^*)]^T \boldsymbol{\lambda} + [\nabla \mathbf{g}(\mathbf{x}^*)]^T \boldsymbol{\mu} &= 0 \quad (n \text{ equations}) \\ \mathbf{h}(\mathbf{x}^*) &= 0 \quad (m \text{ equations}) \\ \forall\ i = 1, \ldots, p, \quad \mu_i g_i(\mathbf{x}) &= 0 \quad (p \text{ equations}) \end{aligned}$$

where $\boldsymbol{\lambda}$ are Lagrange multipliers and $\boldsymbol{\mu} \ge 0$ are the Kuhn-Tucker multipliers.

9. **Penalty & Barrier Functions**

To minimise $f(\mathbf{x})$ subject to p inequality constraints $g_i(\mathbf{x}) \leq 0$, $i = 1, \ldots, p$, define

$$q(\mathbf{x}, p_k) = f(\mathbf{x}) + p_k P(\mathbf{x})$$

where $P(\mathbf{x})$ is a penalty function, e.g.

$$P(\mathbf{x}) = \sum_{i=1}^{p} (\max[0, g_i(\mathbf{x})])^2$$

or alternatively

$$q(\mathbf{x}, p_k) = f(\mathbf{x}) - \frac{1}{p_k} B(\mathbf{x})$$

where $B(\mathbf{x})$ is a barrier function, e.g.

$$B(\mathbf{x}) = \sum_{i=1}^{p} \frac{1}{g_i(\mathbf{x})}$$

Then for successive $k = 1, 2, \ldots$ and p_k such that $p_k > 0$ and $p_{k+1} > p_k$, solve the problem

minimise $q(\mathbf{x}, p_k)$